



Robust and fast pedestrian detection method for far-infrared automotive driving assistance systems



Qiong Liu^{a,b,*}, Jiajun Zhuang^{a,b}, Jun Ma^b

^a School of Software Engineering, South China University of Technology, Guangzhou 510006, China

^b School of Computer Science and Technology, South China University of Technology, Guangzhou 510006, China

HIGHLIGHTS

- A robust and fast far-infrared automotive pedestrian detection method is presented.
- Estimate potential pedestrian regions using pixel-gradient oriented vertical projection.
- PEWHOG is more effective for far-infrared pedestrian representations.
- Iteratively training procedure is presented to generate more robust classifier.
- Experimental results indicate the presented method is effective and promising.

ARTICLE INFO

Article history:

Received 23 October 2012

Available online 20 June 2013

Keywords:

Pedestrian detection

Far-infrared

ROIs generation

HOG

SVM

Automotive driving assistance systems

ABSTRACT

Despite considerable effort has been contributed to night-time pedestrian detection for automotive driving assistance systems recent years, robust and real-time pedestrian detection is by no means a trivial task and is still underway due to the moving cameras, uncontrolled outdoor environments, wide range of possible pedestrian presentations and the stringent performance criteria for automotive applications. This paper presents an alternative night-time pedestrian detection method using monocular far-infrared (FIR) camera, which includes two modules (regions of interest (ROIs) generation and pedestrian recognition) in a cascade fashion. Pixel-gradient oriented vertical projection is first proposed to estimate the vertical image stripes that might contain pedestrians, and then local thresholding image segmentation is adopted to generate ROIs more accurately within the estimated vertical stripes. A novel descriptor called PEWHOG (pyramid entropy weighted histograms of oriented gradients) is proposed to represent FIR pedestrians in recognition module. Specifically, PEWHOG is used to capture both the local object shape described by the entropy weighted distribution of oriented gradient histograms and its pyramid spatial layout. Then PEWHOG is fed to a three-branch structured classifier using support vector machines (SVM) with histogram intersection kernel (HIK). An off-line training procedure combining both the bootstrapping and early-stopping strategy is introduced to generate a more robust classifier by exploiting hard negative samples iteratively. Finally, multi-frame validation is utilized to suppress some transient false positives. Experimental results on FIR video sequences from various scenarios demonstrate that the presented method is effective and promising.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Vision-based automatic pedestrian detection has become a hot spot in recent researches because it is widely used in automotive driving assistant systems [1–3], video surveillance [4,5], content-based image/video retrieval [6], etc. Pedestrian-to-vehicle collision happens more frequently after sundown [7] and the total number

of traffic fatalities involving pedestrian is several times higher at night than daytime [8]. Therefore, it is necessary to explore reliable night-time pedestrian detection component for automotive driving assistant systems. Cameras in visible spectrum can be strongly influenced by illumination conditions and relatively infrared ones are more suitable to capture information at night, especially far-infrared (FIR) cameras because no active illumination is required. The progress for night-time pedestrian detection is also promoted due to the decreasing cost of FIR cameras in recent years.

Vision-based pedestrian detection method generally consists of two main modules: regions of interest (ROIs) generation and pedestrian recognition. We focus on the advances of monocular

* Corresponding author at: School of Software Engineering, South China University of Technology, Guangzhou 510006, China. Tel.: +86 13378680689; fax: +86 02039380258.

E-mail addresses: liuqiong@scut.edu.cn (Q. Liu), zhuang.jj@mail.scut.edu.cn (J. Zhuang), majun0614@163.com (J. Ma).

pedestrian detection that share the insights for night-time automotive applications. For more comprehensive survey of recent works on pedestrian detection, the reader is referred to the work of Gerónimo et al. [9] and Dollár et al. [10].

The popular sliding window approach [11] that generates ROIs over multiple scales from the input images is not suitable for real-time automotive pedestrian detection. Inspired by sliding window approach, Sun et al. [12] presented a keypoint-centric based local sliding window technique for ROIs generation. In their method, all the candidate keypoints in images were detected using SUSAN detector and then ROIs were generated within the neighborhood of the detected keypoints. However, pedestrians are usually warmer and hence appear as brighter objects in FIR images from a local perspective and this special clue was not exploited in [12]. Global image thresholding segmentation was presented to generate ROIs by Bertozzi et al. [13], where the threshold was derived from the statistical properties of pre-collected images containing only background targets since the pixels from pedestrians were much brighter than that from background targets in their dataset. But the global thresholding segmentation technique faces difficulty in handling some potential difference in appearance of pedestrians among different image frames because pedestrians may not always be brighter than background from a global perspective. To address this problem, Ge et al. [2] proposed an adaptive local dual-thresholding segmentation algorithm to generate ROIs. But their dense handling within the whole input images is computationally intensive and might generate too many negative ROIs. Unlike the conventional region-growing used in [14], ROIs were generated using feature-based region-growing with high intensity seeds [1] and the algorithm stops when the connected regions' enclosing bounding boxes no longer cover the possible intervals of pedestrians' aspect ratios. Alternatively, Fang et al. [15] estimated ROIs' horizontal location using intensity-based horizontal projection and determined their vertical location through intensity/body-line-based vertical segmentation. Li et al. [4] subsequently proposed a similar method by combining both intensity-based horizontal and vertical projections. The intensity-oriented projection method is flexible, but the accuracy of generated ROIs heavily relies on the quality of infrared images since it requires that intensity of pixels from pedestrians should be higher than the average pixel intensity of the whole input images in a global view.

When a group of ROIs is generated, further validation will be performed in pedestrian recognition module. Following a learning-based discriminative framework, exploring more discriminative descriptors for pedestrian representations and designing more powerful learning algorithms have always been the pursuits. The discriminative descriptors include Haar wavelets [2,16,17], local binary/ternary patterns (LBP/LTP) and their variants [12,18,19], shapelet [20], edgelet [11,21], intensity self similarity (ISS) [22], histograms of oriented gradients (HOG) and its variants [23–26] etc. Then the extracted features are fed to different classifiers. Support vector machines (SVMs) and boosting algorithms are the two occupied learning algorithms and show excellent performance over state-of-the-art methods in pedestrian recognition [9].

One of the first pioneering efforts was the work of Papageorgiou and Poggio [17] where the combination of Haar wavelets and a polynomial SVM was presented. Inspired by the Haar wavelets, Viola et al. [16] introduced Haar-like features for pedestrian representations and proposed a cascade AdaBoost learning framework for both automatic feature selection and efficient pedestrian detection. Ge et al. [2] extended the cascade detection framework and introduced a two-stage (cascade) tree-structure near-infrared pedestrian classifier using Gentle-AdaBoost. Similar tree-structure detection framework was also proven to be effective by Xu et al. [27]. The dense HOG descriptor was designed specifically for pedestrian representations by Dalal and Triggs [23], and has since

showed excellent performance for finding pedestrians in visible spectrum. Zhu et al. [24] introduced integral histograms to extract HOG for use in cascade-of-rejectors and they observed that the more informative HOG blocks are those locating in local edge or contour regions of pedestrians. In later work, O'Malley et al. [1] successfully extended HOG to pedestrian recognition on automotive FIR videos. Sun et al. [12] extended LBP using the spatial layout of texture cells to describe the symmetrical characteristic of FIR pedestrians and proposed pyramid binary pattern (PBP). And PBP was performed effectively with an SVM classifier. Inspired by the descriptor for characterizing color self similarity (CSS) [28] in the visible spectrum, Miron et al. [22] proposed ISS based on the relative intensity self similarity within specific FIR pedestrian regions, e.g. the head region of pedestrians shares more similar gray-level intensity.

Despite considerable descriptors have been proposed, those based on image gradients (e.g. HOG-like descriptors) are still probably most effective for pedestrian detection [9,10]. Although the work in [19,29,30] demonstrate that the idea of fusing different descriptors can benefit better recognition performance, we only focus on exploring more discriminative single descriptor in this work, because it can also guarantee superior performance when fusing with other descriptors. By further exploiting the underlying data characteristic of FIR images and following the conclusions of [24], we explore the potential of dense HOG descriptor for FIR pedestrian detection by introducing the entropy of distribution of oriented gradient histograms. Gao [31] presented a close procedure, where the entropy of HOG was calculated to represent the image texture and then entropy thresholding was used to directly filter out the regions with either dense texture or textureless that were regarded as ambiguous for image matching and registration. However, it does not necessarily mean that complete or over-complete feature set is redundant according to the idea of dense HOG. This indicates that the performance of pedestrian recognition would decrease if we filter out the features estimated as less informative directly. In addition, we also consider that the training procedure (often underestimated in literature) is crucial, because the recognition performance usually depends on the adopted training procedure when the initial training data and learning algorithms are fixed [23,28].

This work focuses on developing a robust and efficient night-time pedestrian detection system for automotive applications based on monocular FIR camera. A novel recognition framework is proposed to learn and recognize FIR pedestrians, by taking the advantage of an effective descriptor termed pyramid entropy weighted histograms of oriented gradients (PEWHOG) and a new training procedure. The main contributions of this paper are as follows: (1) Pixel-gradient oriented vertical projection is proposed to estimate the input FIR image and reduce its searching regions efficiently. It can be regarded as a preliminary segmentation procedure with which other accurate ROIs generation techniques (e.g. thresholding segmentation) can be coupled. The combination can generate nearly the same positive ROIs and much fewer negative ROIs with a lower runtime. (2) Novel PEWHOG is proposed for more effective FIR pedestrian representations by capturing both the local object shape using the entropy weighted distribution of oriented gradient histograms and its pyramid spatial layout. Then a three-branch structured classifier based on PEWHOG and SVM is used to address the high within-class variance problem caused by pedestrians' imaging size. (3) An iteratively training procedure combining both the bootstrapping and early-stopping strategy is proposed to generate a more robust classifier with lower prediction error. (4) Extensive experiments including both classifier-level and system-level performance evaluations have been conducted to verify the effectiveness of the presented method, and the results demonstrate that it is robust and suitable for real-time applications.

The remainder of this paper is organized as follows: We describe the process of ROIs generation in Section 2 and present the methodology of our pedestrian recognition module in Section 3. Experiments are shown in Section 4, and conclusions are drawn in Section 5.

2. ROIs generation

In order to both guarantee a fast ROIs generation process and suppress as many resulting negative ROIs as possible, we propose an efficient and accurate ROIs generation scheme. It first automatically estimates all the possible vertical image stripes from the input FIR images and filters out those stripes without significant pixel-intensity change using pixel-gradient oriented vertical projection (also called pre-segmentation hereinafter). Then, it generates ROIs more accurately within the estimated vertical image stripes using image thresholding segmentation.

2.1. Pixel-gradient oriented vertical projection

Novel pixel-gradient oriented vertical projection is proposed to effectively filter out the vertical image stripes that do not contain any pedestrian. Gradient vertical projection curve is first defined as the number of pixels with high gradient magnitude versus their corresponding horizontal positions in the gradient image. The curve can generally be partitioned into several stripes with left rising points and right falling points (turning points). In an image stripe, the turning points represent the pixels whose intensity changes either from darkness to brightness or from brightness to darkness more rapidly. The steps of pixel-gradient oriented vertical projection are as follows:

- (1) Compute the gradient image. The mask used to calculate the gradient is 1-D centered point derivatives, e.g. $[-1, 0, 1]$ for calculating image gradient along the horizontal axis.
- (2) Image gradient depicts the information around the boundaries of objects. A threshold T_g is introduced to discriminate whether the pixels belong to the boundaries or not, because small gradient magnitude usually indicates that the adjacent pixels share similar intensity or are homogeneous. Then the pixels from the boundaries are preserved and some interference is suppressed. T_g should be selected carefully to ensure that the images columns containing pedestrians will have nonzero projection in the gradient vertical projection curve. An example of resulting binary gradient image is shown in Fig. 1b. Then the gradient vertical projection curve can be generated from the binary gradient image, as shown in Fig. 1c.
- (3) Automatically search all the waves with left rising points and right falling points from the projection curve. Then segment the input image into several vertical stripes by pairing left rising points and right falling points.

Different from intensity-based projection methods presented in [4,15] where all the turning points of the projection curve are recorded, only the turning points whose value is higher than a threshold T_s are considered in this work. Because pedestrians usually feature more pixels with evident gradients in the boundaries than many background regions, we can ignore some image strips containing fewer pixels in this way. T_s is automatically determined by statistical property of the projection curve with the form:

$$T_s = \omega \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)^2} \quad (1)$$

where x_i is the number of pixels in i th column of the projection curve, $\mu_x = \frac{1}{n} \sum_{i=1}^n x_i$, n is the length of the columns and ω is a weight coefficient. Moreover, the width (intervals between the two paired turning points) of the vertical image stripes passing through pedestrian regions should satisfy a minimum value, e.g. the minimum width of possible pedestrian samples, so it is reasonable that those stripes with smaller width are also filtered out.

Based on the above steps, the searching regions of the whole input image can be reduced to several highlighted vertical stripes, some of which contain pedestrians, as shown in Fig. 1d. The dotted-line in Fig. 1c indicates the resulting T_s . The ground and sky in FIR images usually represent as large homogeneous regions. This characteristic makes it easier and more reliable to perform pixel-gradient oriented vertical projection using the gradient information.

Unlike the vertical image stripes that contain some certain amount of both homogeneous ground and sky regions, horizontal image stripes passing through pedestrians usually contain too many unexpected image artifacts or objects with various intensity distributions, which can affect the completeness of segmented pedestrians. As a consequence, pixel-gradient oriented horizontal projection performs poorer for estimating pedestrian regions and hence is not considered in this work.

2.2. Image segmentation and ROIs selection

Image thresholding is a common and simple segmentation technique to extract foreground objects from FIR images. Although FIR pedestrians are generally represented as brighter objects, the overall brightness of pedestrians may not always be uniform from a global perspective. Fortunately, it still holds a more realistic assumption that the pixels from infrared pedestrians are usually brighter than the nearby pixels on both adjacent sides of pedestrians from each horizontal scan line [2]. Consequently, we adopt the local dual-thresholding segmentation presented in [2] to perform image segmentation within the estimated image stripes.

In order to filter out the noise in the resulting binary segmentation image and compensate for the weak-connected regions, morphological erosion operation with a horizontal mask of 3-by-1 and dilation operation with a square mask of 3-by-3 are performed successively on the resulting binary image. The final binary segmentation result is shown in Fig. 1e, in which all the connected regions (i.e. ROIs) are marked with pink rectangles. According to the fact that the aspect ratios of pedestrians usually follow some certain distribution, some of ROIs in the binary image can further be filtered out before being passed to the pedestrian recognition module. An example of all the selected ROIs (marked with blue¹ rectangles) using the above tricks is illustrated in Fig. 1f.

Although image thresholding segmentation is adopted in this paper based on the estimated image stripes, other potential techniques for generating ROIs more accurately may also be coupled with the pixel-gradient oriented vertical projection. Such combination provides faster implementation versions for the certain techniques, since it can guarantee nearly the same positive ROIs generation results with a much lower runtime, as well as with a much smaller number of generated negative ROIs (because most image stripes containing only background are filtered out).

3. Pedestrian recognition

Feature representations are no doubt the first critical factor when designing a pedestrian classifier. In this work, novel EWHOG

¹ For interpretation of color in Figs. 1 and 9, the reader is referred to the web version of this article.

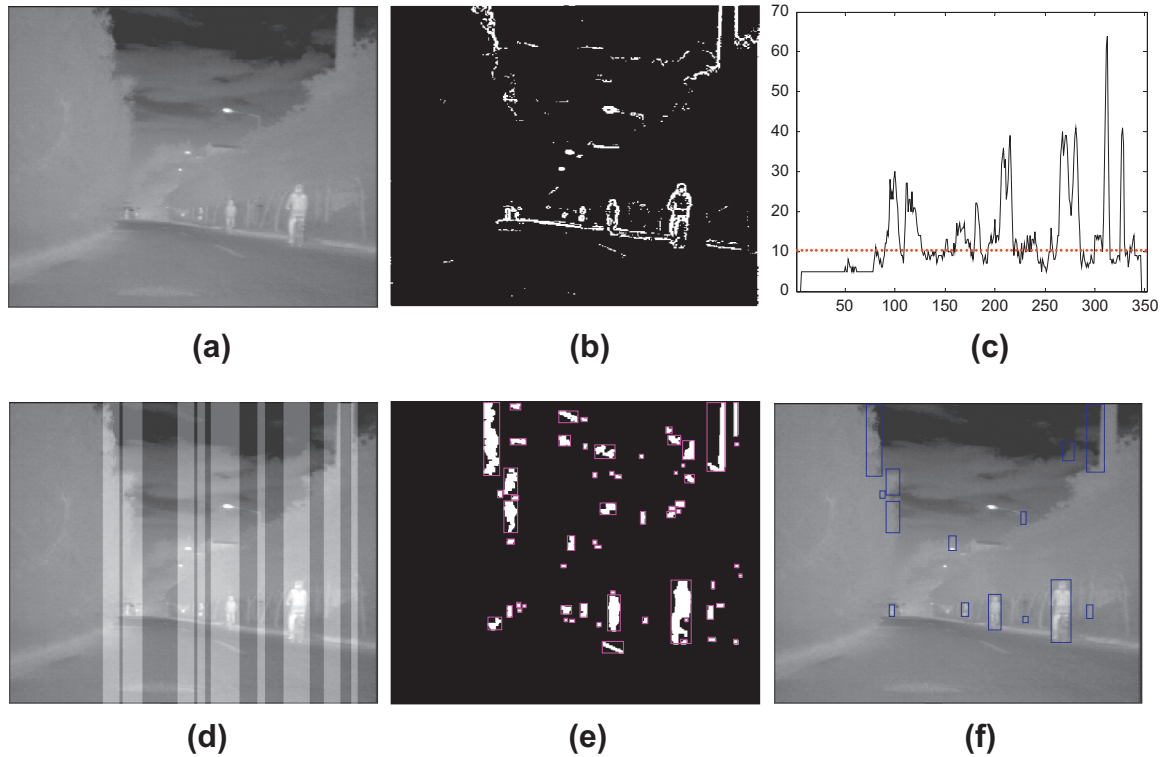


Fig. 1. Example of ROIs generation: (a) raw image, (b) binary gradient image, (c) gradient vertical projection curve, (d) estimated vertical image stripes (highlighted ones), (e) image segmentation result based on the estimated vertical image stripes and (f) selected ROIs.

(entropy weighted histograms of oriented gradients) at multiple levels in spatial pyramid space is proposed to describe FIR pedestrians. The scheme of our pedestrian detection is shown in Fig. 2. During the off-line training phase, SVM using histogram intersection kernel (HIK), together with a new iteratively training procedure, are proposed to generate the pedestrian classifier. During the on-line detection phase, multi-frame validation is introduced to suppress some false positives.

3.1. Entropy weighted histograms of oriented gradients (EWHOG)

HOG proposed by Dalal and Triggs [23] is a very popular descriptor for pedestrian detection. The basic idea of HOG is that local object shape can be characterized rather well by the distribution of local intensity gradients or edge directions in a dense fashion. And as demonstrated in [24], the most informative components (achieve a higher recognition accuracy) carried by HOG are those extracted from the local edge or contour regions of pedestrians, while the inner texture of pedestrians is generally not considered and may be less informative. A common observation is that most pedestrians or background targets in FIR images are usually lack of rich texture but with relative notable local shape or edge. As a consequence, it is beneficial to exploit the underlying data characteristic of FIR pedestrians in the procedure of extracting HOG. To this end, a novel variant of dense HOG (i.e. EWHOG) is proposed to describe FIR pedestrians.

The basic idea is that the components of HOG extracted from local edge regions of pedestrians can be weighted more than those from the inner regions of pedestrians. According to the data characteristic of FIR pedestrians, we found that the local edge regions usually represent as more heterogeneous because the pixels in these regions always belong to different classes (e.g. pedestrian or non-pedestrians); while comparative uniform ones usually indicate that they probably locate in the inner regions. Entropy is

widely used to estimate the systematic microdistributed chaotic degrees in information theory and the higher the entropy is, the more informative it is. Inspired by this concept, entropy is introduced to estimate the gradients distributed degrees of different regions, e.g. various blocks in dense HOG. In other words, HOG extracted from different regions can be weighted differently. In our case, the k th block carries a weight W_k :

$$W_k = -\sum_{t=1}^{c \times r} P_t \log_2 P_t \quad (2)$$

$$P_m = h_m / \sum_{t=1}^{c \times r} h_t, m = 1, 2, \dots, c \times r \quad (3)$$

where h_t is the strength of the t th histograms of oriented gradients in k th block, c is the number of cells within each block and r is the number of orientation bins in each cell. W_k is calculated before block normalization because the normalization may alter the intrinsic distribution of oriented gradient histograms. Although the trilinear interpolation in HOG might avoid generating a sparse feature vector, we define:

$$P_n \log_2 P_n = 0 \quad (4)$$

where P_n is equal to zero.

397 positive (pedestrian) and 458 negative (non-pedestrian) samples are collected to investigate the distribution of entropy carried by different blocks of FIR samples. The distribution of the average entropy is shown in Fig. 3. All the samples are normalized into a spatial resolution of 32×80 pixels, and other settings used to extract EWHOG are as follows:

- Size of each cell: 4×8 pixels.
- Size of each block: 2×2 cells.
- Overlapping area between two adjacent blocks: 1 cell size.

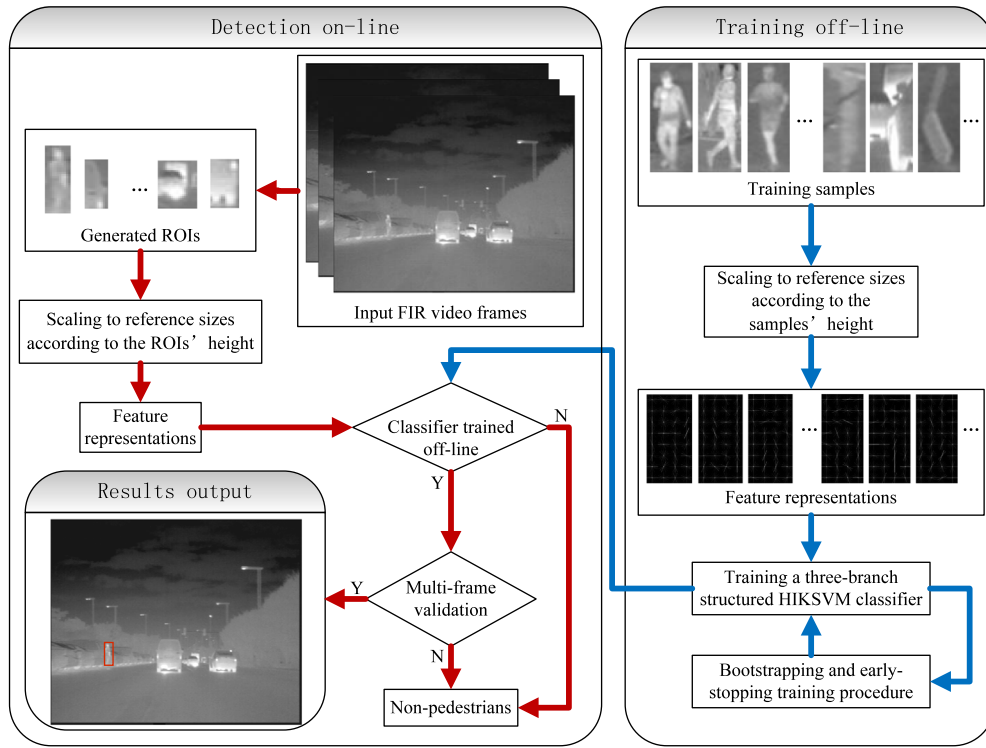


Fig. 2. Scheme of our pedestrian detection.

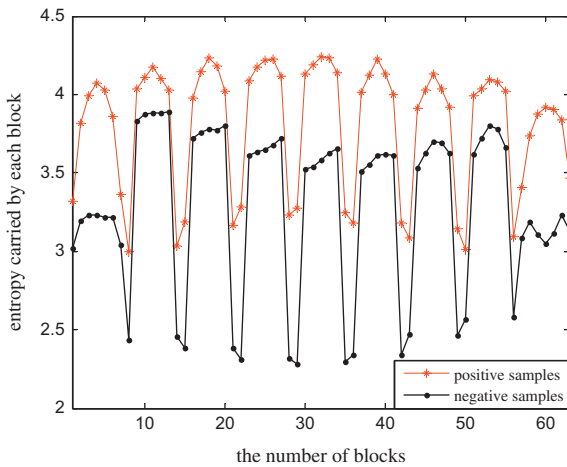


Fig. 3. Distribution of average entropy carried by positive and negative samples.

- Number of orientation bins in each cell: 9 bins over 0–180°.

The blocks located in the local edge regions usually carry higher entropy than those located in the inner regions for a certain class of FIR samples, e.g. pedestrians. The reason is that the heterogeneous local edge regions could generate more uniform distribution of oriented gradient histograms while homogeneous inner regions usually generate sparser distribution of gradient histograms with some certain orientations. It also demonstrates that the entropy carried by local edge regions of positive samples is generally higher than that of negative samples according to Fig. 3, because non-rigid pedestrians usually share more abundant/irregular shapes and their local edge regions often carry higher entropy. On the contrary, the negative samples often share certain types of some regular shapes, e.g. the strictly vertical edge of a lamppost, and such certainty information is bound to result in lower entropy.

Consequently, the entropy carried by each block can be used to improve the discriminative power of dense HOG for FIR pedestrian detection. EWHOG is proposed by weighting the resulting entropy on the corresponding histograms of oriented gradients extracted from each block after block normalization, so as to pay more emphasis on the distribution of local intensity gradients provided by local object shape. EWHOG is a median-level descriptor because local chaotic degree of oriented gradient histograms is further extracted from the original image gradients.

3.2. Pyramid EWHOG

In order to capture both the local shape distribution of pedestrians and its spatial layout for more effective FIR pedestrian representations, pyramid entropy weighted histograms of oriented gradients (PEWHOG) is further proposed. Specifically, spatial layout of local cells partition is introduced into EWHOG using a spatial pyramid image representation fashion. As illustrated in Fig. 4c–e, by doubling the number of pixels repeatedly in both axis directions, each pedestrian image is partitioned into a group of increasingly finer spatial cells. In our case, the cells partition manner at each level of resolution represents the distribution of oriented gradient histograms at that level. A EWHOG feature vector is extracted at each pyramid resolution level of local cells partition and the Final PEWHOG feature vector is the concatenation of all the EWHOG feature vectors extracted at different levels. In this way, complementary feature sets can be generated.

In the pyramid pedestrian representation, there are $2^l \times 2^{l-1} = 2^{2l-1}$ cells at level l ($l \in [1, L]$). Assume that each cell is represented by an r -dimension feature vector with respect to the r orientation bins and 4 (2×2) cells form a block. There is a special case that each block contains only one cell at level 1 as only one cell is generated along the horizontal axis. As a consequence, level l generates a $r \times 2^{2l-1}$ -dimension feature vector and the dimensionality of PEWHOG feature vector is $r \sum_{l=1}^L 2^{2l-1}$, if we do not consider the case of blocks overlapping in dense EWHOG.

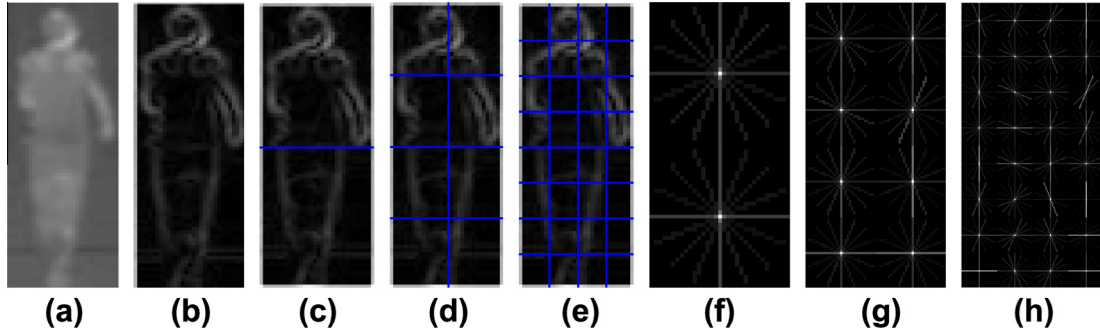


Fig. 4. Example of spatial pyramid representation of an FIR pedestrian using PEWHOG: (a) raw image, (b) gradient image of (a), (c–e) cells partition using pyramid-fashion representation from level $l = 1$ to $l = 3$, respectively, (f–h) resulting feature representations showing the oriented gradient histograms extracted in each cell from level $l = 1$ to $l = 3$, respectively.

Fig. 4 schematically illustrates the PEWHOG coding scheme, showing an FIR pedestrian image and its corresponding intuitive PEWHOG representation with pyramid level $L = 3$. The pedestrian image is scaled into a spatial resolution of 32×80 pixels in the example. PEWHOG is different from the scale space pyramid representation of local shape distribution in an image because no smoothing among the levels of the pyramid representation is required, and instead the local shape distribution is extracted from the original (scaled) resolution image.

3.3. Three-branch structured support vector machines (SVM) classifier

Being good at dealing with the non-linear and small sample pattern classification issues, SVM has become one of the most popular classifiers in the field of object detection and is adopted in the pedestrian recognition module in this work.

Given N labeled training samples of the form $\{(x_i, y_i)\}_{i=1}^N$, with $y_i \in \{+1, -1\}$ and $x_i \in \mathbb{R}^d$, the aim of SVM is to find an optimal hyper-plane which best separates the two classes of samples by minimizing the objective function:

$$F(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N \xi_i, C \geq 0 \quad (5)$$

$$s.t. \ y_i(\omega^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

The optimal solution of the above problem is usually obtained by maximizing the dual form of Eq. (5):

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (6)$$

$$s.t. \ \sum \alpha_i y_i = 0, 0 \leq \alpha_i \leq C$$

The decision function is $sign(h(x))$, where:

$$h(x) = \sum_{i=1}^v \alpha_i y_i K(x, x_i) + b \quad (7)$$

where v is the number of resulting support vectors. Because PEWHOG is a type of histogram features and HIK is appropriate to evaluate the similarity between two histogram features [32], HIK is introduced as the kernel function for SVM classification with the form:

$$K(x, z) = \sum_{i=1}^d \min(x(i), z(i)) \quad (8)$$

3.3.1. Structure of the classifier

The imaging size of pedestrian with various distances to the camera can vary from each other significantly due to the moving

platform, which may further aggravate the appearance variability of pedestrian presentations. We divide the original training samples into several disjoint subsets and train a multi-branch structured classifier in order to reduce the influence caused by the frequently changing imaging size and improve the distribution compactness of positive samples in the feature space. The width of pedestrians might change dramatically due to various postures even at the same distance to the camera. E.g. for pedestrians with similar height (physical height of a pedestrian), the width of pedestrians with outstretched hands is usually much larger than that of those with hands hanging down. Fortunately, the height (imaging height of a pedestrian) of them differs little in this case. Therefore, the original training samples are divided into three disjoint subsets according to the distribution of samples' height: the samples more than 64 pixels in height are categorized as near targets and scaled into a spatial resolution of 32×80 pixels, those less than 32 pixels in height are categorized as far targets and scaled into a spatial resolution of 12×32 pixels, and the rest (as median targets) are scaled into a spatial resolution of 24×64 pixels. Then three PEWHOG-based HIKSVM sub-classifiers are trained independently and form a three-branch structured classifier for pedestrian recognition. Our previous work [33] has already demonstrated the effectiveness of the classification structure.

3.3.2. Training procedure

Pedestrian detection is usually treated as a binary classification task and rare pedestrians are distinguished from enormous background targets. In such a rare-event-detection issue, the specific positive samples can be collected (relatively) easier than the negative samples to some extent since it is not a trivial task to collect an enough representative dataset that can well cover the tremendous possible negative patterns from the real world scenarios. It is necessary to search representative patterns for the negative dataset during the off-line training phase. Besides, Walk et al. [28] indicated that selecting too many initial negative samples arbitrarily might not improve the overall detection accuracy significantly in most cases. This can also be explained that the initial negative samples in the training data seem to be not representative enough. However, additional bootstrapping iterations can force the classifier to concentrate on the more representative (hard) samples and generate a more robust classifier because one can imagine that a classification mistake can reinforce the classifier itself.

In this work, bootstrapping and early-stopping strategy [34] are introduced to search hard negative samples iteratively. Specifically, the trained classifier in last iteration uses its own predictions to search hard negative samples from the video sequences containing background targets, and the generated hard samples is then added to the initial training data to re-train the classifier. The

procedure repeats until the overall performance of the re-trained classifier on the test video sequences no longer improves. The video sequences used to generate hard negative samples and the test video sequences are disjoint from each other. The training procedure also deals with the disadvantage that the recognition performance depends heavily on the initial training data (one may not collect an appropriate training dataset easily), because the influence caused by using only initial training data will decrease after several iterations, which confirms the strategy to concentrate more on representative samples.

3.4. Multi-frame validation

We observe that pedestrians can usually be correctly detected in several successive video frames (e.g. 5 frames), while many false positives usually occur for a smaller number of successive frames or probably flash only in one single frame. In such case, the introduction of multi-frame validation strategy can readily suppress many false positives almost without affecting the true positives.

4. Experiments

In this section, we conduct experiments to answer several questions: (1) How do the parameter settings in pixel-gradient oriented vertical projection affect the pre-segmentation results? (2) What improvement on recognition performance does PEWHOG provide compared to conventional HOG-like descriptors? (3) What benefit does the proposed training procedure offer with respect to the recognition accuracy? (4) Does the system-level performance of the integrated pedestrian detection system meet the criteria for automotive applications, i.e. both detection accuracy and processing speed? For the above purposes, on-road video data (with a resolution of 352×288 pixels) was collected using an FIR camera mounted on a car at driving speed no more than 60 km/h. The video data was captured in both winter and early summer scenarios in Guangzhou City, China. The FIR camera is of the wavelength band 8–14 μm with a frame rate of 25 frames per second (f/s). All the experiments are performed on an Intel Pentium E5800 (3.20 GHz) workstation with 2 GB RAM using MATLAB.

A database consisting of gray scale image cut-outs (samples) is collected from the on-road video data. About half of the samples in the database are cropped from the video data manually, and the rest are generated using the proposed ROIs generation module. Some examples are shown in Fig. 5.

4 Test video sequences are extracted from the on-road video data and pedestrians in the video sequences are annotated with bounding boxes (ground truth). Because it is quite difficult to annotate distant pedestrians (too far away), even for human operators attempting to determine the ground truth, the distant pedestrians less than 12 pixels in height are not considered. Similar handling is also considered in [10], in which annotations with less than 30 pixels in height are deemed to be ambiguous targets. 2 Video sequences containing mainly background targets are extracted for the proposed training procedure (to generate hard negative samples). Besides, in order to avoid the over-fitting on the test video sequences during the evaluation of the training procedure, 2 additional validation video sequences are prepared. The details about the experimental video sequences are listed in Table 1.

4.1. Evaluation criteria

Since the pre-segmentation aims at reducing the searching regions while preserving the pedestrians as many as possible, image stripes ratio (ISR) and miss rate (MR) are defined to evaluate the performance with respect to different parameter settings (i.e. T_g and T_s) in pixel-gradient oriented vertical projection. ISR is

defined by the ratio between the sum area of all estimated image stripes and the area of the whole input image, while MR is given by:

$$\text{MR} = \frac{n_M}{n_A} \quad (9)$$

where n_M is the number of missed pedestrians and n_A is the total number of annotated pedestrians in test video data.

For the evaluation of classifier-level performance, two metrics, i.e. true positive rate (TPR) and false positive rate (FPR) are defined as follows:

$$\text{TPR} = \frac{\text{number of correctly classified positive samples}}{\text{total number of positive samples}} \quad (10)$$

$$\text{FPR} = \frac{\text{number of incorrectly classified negative samples}}{\text{total number of negative samples}} \quad (11)$$

The huge number of the total negative ROIs is usually worthless for system-level performance evaluation because a large amount of generated easy negative ROIs tend to significantly lower the FPR. A more reasonable false alarm rate (FAR) is defined by:

$$\text{FAR} = \frac{n_{\text{FP}}}{n_f} \quad (12)$$

where n_{FP} is the number of false positives and n_f is the number of image frames in test video data. The definition of FAR indicates the average number of false positives per video frame. The detection rate (DR) is defined by:

$$\text{DR} = \frac{n_D}{n_A} \quad (13)$$

where n_D is the number of correctly detected pedestrians. A positive detection in the video sequences is considered correct only if the overlapping area a between the detected bounding box B_d and the annotated ground truth B_{gt} exceeds 0.35 in this work, where a is defined by:

$$a = \frac{\text{intersection area } (B_d \cap B_{\text{gt}})}{\text{union area } (B_d \cup B_{\text{gt}})} \quad (14)$$

Because both the detection rate and false alarm rate are crucial in any practical object detection system, a metric termed weighted accuracy (WA) is defined to evaluate the overall performance of the proposed training procedure using the form:

$$\text{WA} = p \frac{n_D}{n_A} + (1 - p) \left(1 - \frac{n_{\text{FP}}}{n_f} \right) \quad (15)$$

where p is a weight that satisfies $0 \leq p \leq 1$. WA can be regarded as the balance between DR and FAR.

4.2. Evaluation of the pre-segmentation

The first group of experiments aims to validate the performance of the proposed pre-segmentation procedure with respect to different settings of parameters T_g and T_s , and determine their optimal settings empirically. FIR image sequences with total 971 frames are randomly extracted from the six test and validation video sequences listed in Table 1. The total number of annotated pedestrians in the image sequences is 759. Because T_s is dynamically determined by Eq. (1), in which the resulting x_i is directly affected by T_g , the pre-segmentation performance is first evaluated using diverse settings of T_g and a fixed weight coefficient of $\omega = 1$ in Eq. (1). Fig. 6a presents the trend of performance change with respect to diverse settings of T_g . It demonstrates that MR is less sensitive to lower threshold settings of T_g . However, the number of missed pedestrians increases when T_g gets larger. The reason is that some pedestrians far away from the camera might represent as more blur

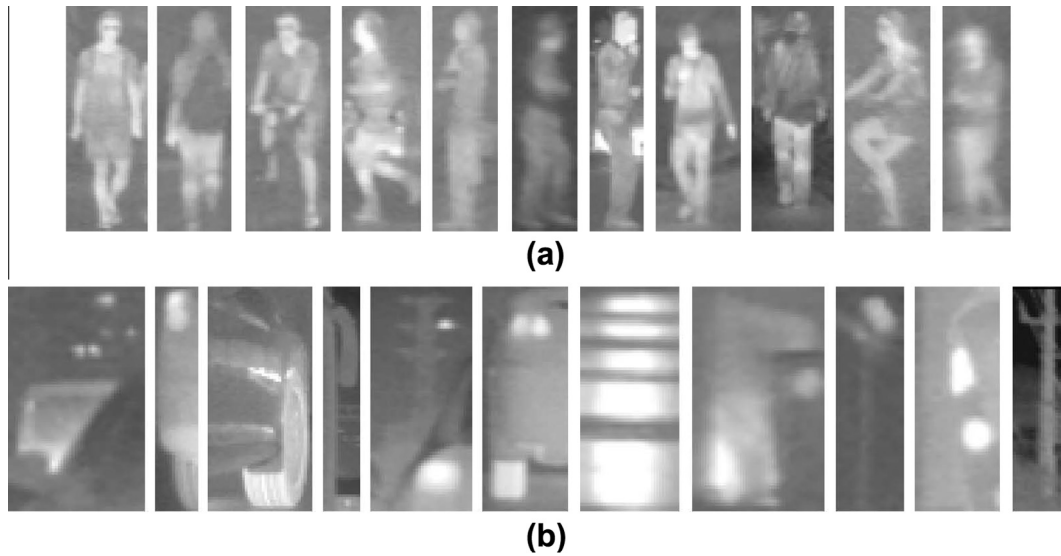


Fig. 5. Some examples from our database: (a) positive samples and (b) negative samples.

Table 1

Details about the experimental video sequences.

Sequence name	Length (total frames)	Annotated pedestrians ^a	Recorded date	Remark
SummerSeq01	1567	494	May 28th 2012	Test video
SummerSeq02	1448	1143	May 28th 2012	Test video
SummerSeq03	824	108	May 28th 2012	Validation video
SummerSeq04	2239	–	May 28th 2012	To generate hard samples
WinterSeq01	2419	1139	November 26th 2011	Test video
WinterSeq02	1553	204	November 26th 2011	Test video
WinterSeq03	1000	966	November 26th 2011	Validation video
WinterSeq04	2208	–	November 26th 2011	To generate hard samples

^a The numeral refers to the number of pedestrians annotated in the video sequence.

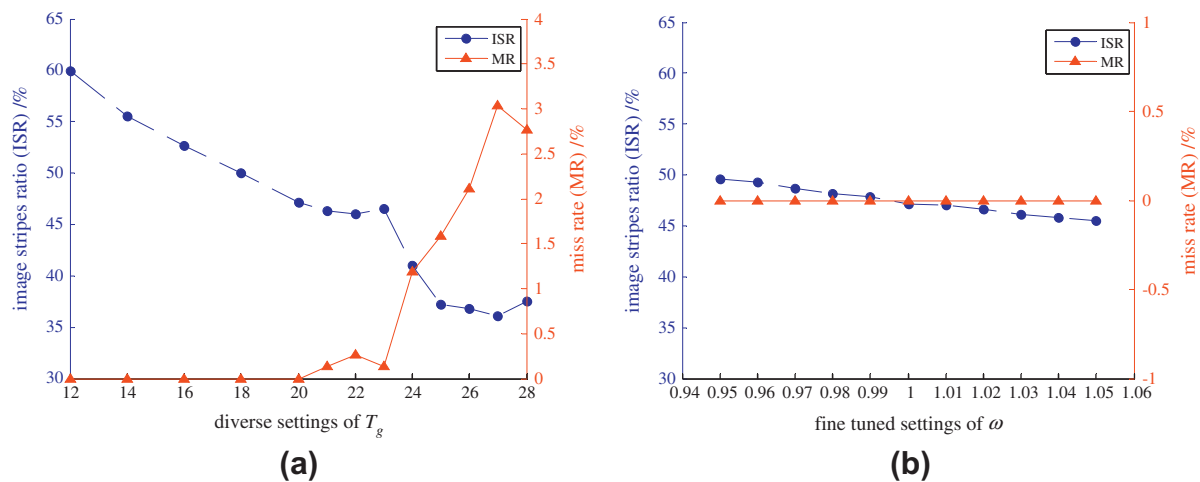


Fig. 6. Evaluation of pre-segmentation: (a) performance using diverse settings of T_g and (b) performance using fine tuned settings of ω .

objects in low resolution FIR images due to the significant reduction of heat radiations during the long-distance transmission. They usually share more similar gray-level intensity with their nearby background and feature with fewer pixels than nearer targets, which results in smaller image gradients around their boundaries. A larger threshold setting of T_g may suppress these image pixels in the binary gradient images and hence the corresponding pixels will not contribute to gradient vertical projection curves. Another primary trend is that ISR significantly decreases against the increasing T_g .

Taking both MR and ISR into account, one can learn that conservative settings of T_g guarantee all the pedestrians could be preserved in the resulting image stripes, but generate larger searching regions. On the other hand, aggressive settings of T_g could obtain smaller searching regions, but might cause some certain missed pedestrians. Consequently, we select $T_g = 20$ for its highest performance when doing the subsequent experiments.

Furthermore, we study the ISR and MR with respect to T_s . As stated above, T_s is dynamically determined by Eq. (1) according to the

statistical property of gradient vertical projection curve and the only parameter that affects T_s is the weight coefficient ω . However, ω is not suitable for significant changes, as it might confuse the statistical property of the projection curves. We fine tune ω from 0.95 to 1.05 for the performance evaluation and the corresponding result is shown in Fig. 6b. The experimental results demonstrate that the performance of pre-segmentation is almost insensitive to T_s , especially MR which is obviously insensitive to different fine tuned settings of ω . The trend of ISR decreases slightly against the minor increase of ω , because larger ω leads to the increasing of T_s and will further influence the selection of turning points. Under the above guidance, we conservatively set $\omega = 1$ for the following experiments.

4.3. Classifier-level performance evaluation

779 Positive samples and 1890 negative samples are selected as the training data from the database. The main settings for extracting PEWHOG are as follows: the number of orientation bins in each cell is 9 over 0–180°, block normalization method is L2-norm and the overlapping area between two adjacent blocks is 1 cell size.

We first study the influence of parameter – pyramid levels L {2,3,4} – on the recognition performance using 10-fold cross validation. The procedure of 10-fold cross validation is repeated for 10 times. Then the average overall accuracy and its standard deviation are calculated to evaluate the recognition performance. In this experiment, all the samples are scaled into a spatial resolution of 32×80 pixels. The average overall accuracy and the corresponding standard deviation under different parameter settings of L {2,3,4} are $90.85\% \pm 7.06 \times 10^{-3}\%$, $93.67\% \pm 7.34 \times 10^{-3}\%$ and $95.48\% \pm 8.67 \times 10^{-3}\%$ respectively. The results demonstrate that higher recognition accuracy is acquired with a larger setting of L because more complementary (over-complete) features can be extracted for pedestrian representations. However, a PEWHOG feature vector with a much higher dimensionality would also be generated in this case, e.g. the dimensionality of the resulting feature vector is nearly 5000 when L equals to 4. It is computationally intensive and can handicap real-time implementation of the integrated detection system. In order to balance the recognition accuracy and computational overhead, L is set to 3.

Then we compare PEWHOG with several conventional descriptors for FIR pedestrian recognition, which include HOG [23], pyramid HOG (PHOG) [31] and PBP [12]. 1060 positive samples and 4067 negative samples are prepared as test data (disjoint with the above training data) for classifier-level performance evaluation. All the training and test samples are divided into three disjoint sub-sets and scaled to reference spatial resolution according to their pixels in height respectively, as stated in Section 3.3.1. Fig. 7 shows the receiver operating characteristics (ROC) curves for the classifier-level performance evaluation using different descriptors.

PBP (and probably other variants of LBP) is generally used for image texture representation and is suitable to detect FIR pedestrians near enough to the camera to some extent. However, the FIR samples usually feature with less or even without rich texture information in our database, especially for distant targets. Therefore, PBP performs poorer than HOG-like descriptors. EWHOG-based classifiers obtain better recognition accuracy than the conventional HOG-based classifiers as we expected. Even without the spatial pyramid representation, EWHOG can also guarantee a competing recognition performance against PHOG. What's more, the introduction of spatial pyramid representation can further improve the recognition performance of EWHOG, which indicates that the proposed PEWHOG is more appropriate and discriminative for FIR pedestrian detection than other descriptors as shown in Fig. 7.

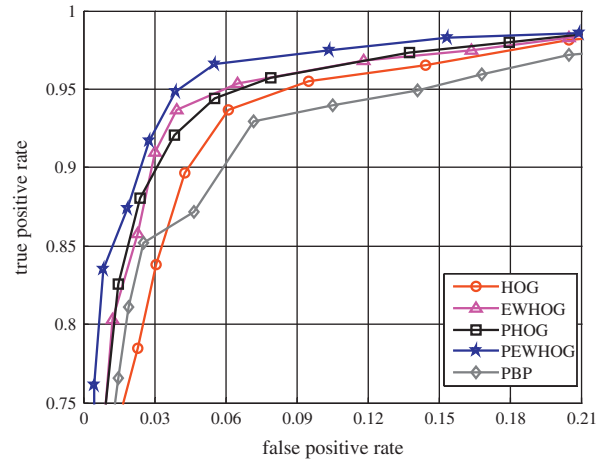


Fig. 7. ROC curves for the classifier-level performance evaluation using different descriptors.

4.4. System-level performance evaluation

According to the proposed training procedure, 1179 positive samples and 788 negative samples are selected from the database randomly as the initial training data, and introduce additional bootstrapping iterations to concentrate more on hard negative samples. The hard negative samples are searched and collected iteratively from the two video sequences SummerSeq04 and WinterSeq04 listed in Table 1. The performance evaluation results on both the test and validation video sequences using the training procedure are illustrated in Fig. 8. The true negative rate (TNR) shown in Fig. 8a is the ratio between the number of correctly classified negative ROIs and total number of generated negative ROIs.

Because the positive samples used in the training procedure are fixed, the decision hyper-plane is shifted a little towards the positive samples while only augment hard negative samples are added into the feature space in each round of iteration, which leads to a little decreasing of DR, as shown in Fig. 8a. Such decreasing of DR is slightly and could be tolerated because much fewer false positives are generated accordingly as shown in Fig. 8b. In this way, the WA is improved. Fig. 8a also indicates that the WA tends to convergence after the fifth iteration, and the procedure of bootstrapping hard negative samples is stopped at this round of iteration according to the early-stopping strategy. Finally, an augment negative dataset with total 2014 samples (initial 788 plus 1226 hard samples) is then used to produce a more robust and generalized pedestrian detector (classifier).

From the perspective of automotive applications, it is reasonable to evaluate the number of detected/missed pedestrians in the video sequences, the average false positives per video frame and the implementation speed of the integrated detection system. The details of the detection results using the 4 test video sequences are listed in Table 2. The integrated detection system runs at an average speed of around 5 frames per second using MATLAB.

Some examples of the detection results are shown in Fig. 9. It should be mentioned that the horizontal viewing angle of the mounted FIR camera is different as shown in the video sequences captured in summer (e.g. Fig. 9a–d) and winter scenarios (e.g. Fig. 9e–g). The former case is regarded as more appropriate for automotive driving assistance systems, as more information about the road usage can be captured in this way. As shown in Fig. 9, pedestrians (or more precisely vulnerable road users) with various types of presentations can be correctly detected, e.g. those standing still on the ground, riding a motorcycle or bicycle, moving across the road, etc. However, missed detections and false positives may also happen. For example, some pedestrians at farther

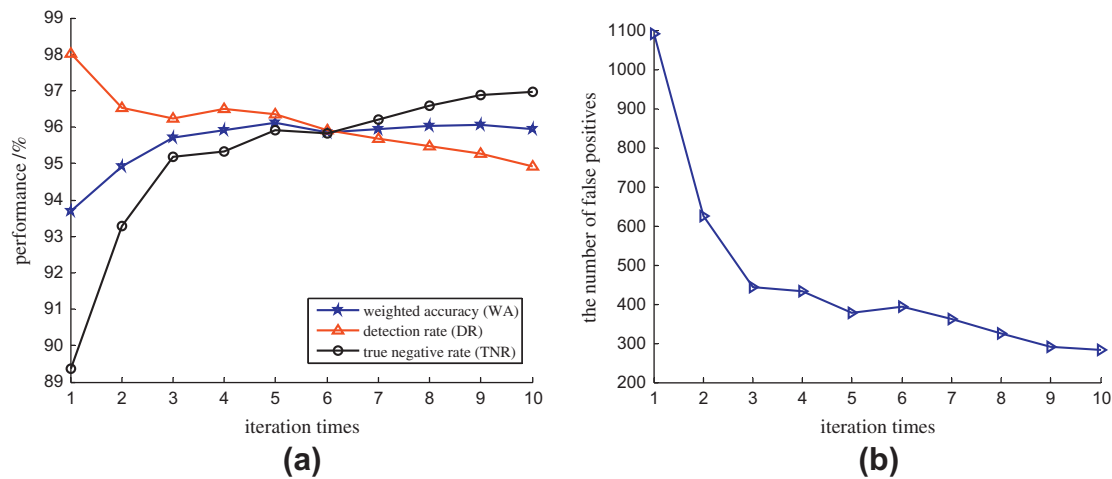


Fig. 8. Performance evaluation using our training procedure: (a) the overall detection performance against iteration times and (b) the total number of false positives against iteration times.

Table 2

System-level performance evaluated on the test video sequences.

Test video	Length (total frames)	Recall ^a	Missed	False positives	DR	FAR
SummerSeq01	1567	487/494	7	40	0.986	0.026
SummerSeq02	1448	1112/1143	31	19	0.973	0.013
WinterSeq01	2419	1101/1139	38	45	0.967	0.019
WinterSeq02	1553	192/204	12	38	0.941	0.024
Average					0.967	0.021

^a The numerator refers to the number of correctly detected pedestrians.

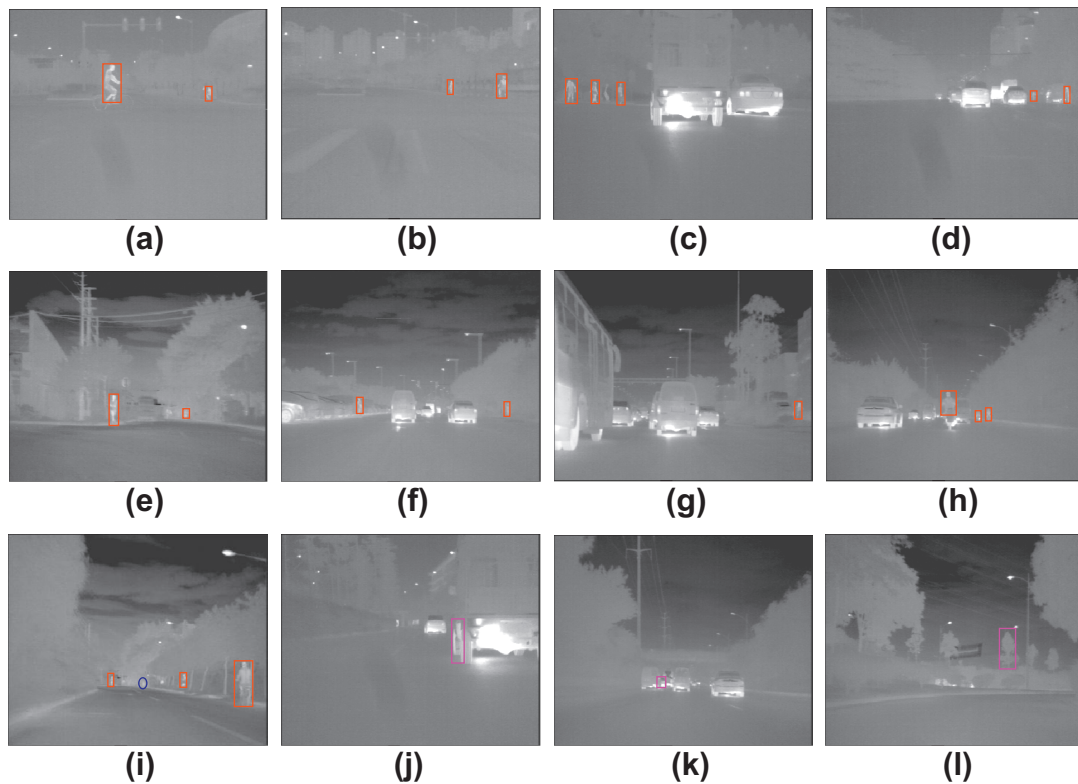


Fig. 9. Some detection examples from the experimental video sequences.

distances are missed, because FIR cameras are generally of low-resolution and pedestrians far away from the camera often feature as ambiguous targets, which usually do not provide sufficient

discriminative information. A missed detection (marked with blue ellipse) illustrating this case is shown in Fig. 9i. Sometimes false positives may be caused by the tyres (or hot rear parts) of driving

vehicles and trees from specific view angles. Some examples of false positives (marked with pink rectangles) are shown in Fig. 9j–l. From the perspective of the practical applications, some false positives (e.g. the one illustrated in Fig. 9l) can be further suppressed using the position constrain, i.e. pedestrians should be on the ground, not hanging in the sky.

The presented pedestrian detection method has been implemented on an experimental vehicle equipped with monocular FIR camera (device model: NV628) as a part of automotive driving assistance systems of Guangzhou SAT Infrared Technology Co., Ltd., Guangzhou, China.

5. Conclusions

This paper presents a night-time pedestrian detection method for automotive driving assistance systems using monocular FIR camera. The major contributions are threefold: (1) an efficient pre-segmentation using pixel-gradient oriented vertical projection is proposed to reduce the searching regions of input FIR images. The method guarantees that almost all the pedestrians can be preserved in the estimated vertical image stripes while most background with large homogeneous regions can be filtered out. Consequently, it speeds up the whole ROIs generation phase and undoubtedly some false positives can also be suppressed since much fewer negative ROIs are generated within the estimated image stripes. (2) In order to describe FIR pedestrians more effectively, a novel descriptor called PEWHOG is proposed to capture both the local object shape using the entropy weighted distribution of oriented gradient histograms and its spatial layout. To deal with the high within-class variance problem caused by pedestrians' imaging size, a three-branch structured classifier using HKSVM is then introduced. Extensive classifier-level experiments have been conducted using image cut-outs from on-road FIR video data and the results indicate that PEWHOG is a more discriminative FIR pedestrian descriptor compared to several conventional descriptors. (3) Regarding the rare-event-detection issue implicit in pedestrian detection, a training procedure combining both the bootstrapping and early-stopping strategy is proposed to generate a more robust classifier by exploiting the more representative negative samples iteratively. The presented pedestrian detection method has been tested on automotive video sequences captured in various scenarios. The system-level experimental results demonstrate that the method is robust and suitable for real-time applications.

There are several issues for further investigation regarding this work. Although a reference range of parameter settings is proven to be effective for pixel-gradient oriented vertical projection procedure through empirical studies, there is still no principled rule to guide the selection of optimal parameter settings. In our future work, we will try to address this problem and further improve its adaptiveness. Besides, even though the training procedure is proposed for bootstrapping hard negative samples, it also deserves to be applied to concentrate on more representative positive samples.

Acknowledgement

The authors acknowledge the financial support of the National Natural Science Foundation of China through Project 61171141.

References

- [1] R. O'Malley, E. Jones, M. Glavin, Detection of pedestrians in far-infrared automotive night vision using region-growing and clothing distortion compensation, *Infrared Physics & Technology* 53 (6) (2010) 439–449.
- [2] J.F. Ge, Y.P. Luo, G.M. Tei, Real-time pedestrian detection and tracking at nighttime for driver-assistance systems, *IEEE Transactions on Intelligent Transportation Systems* 10 (2) (2009) 283–298.
- [3] D. Olmeda, A. de la Escalera, J.M. Armingol, Contrast invariant features for human detection in far infrared images, in: *Proceedings of IEEE Intelligent Vehicles Symposium*, 2012, pp. 117–122.
- [4] J.F. Li, W.G. Gong, W.H. Li, X.Y. Liu, Robust pedestrian detection in thermal infrared imagery using the wavelet transform, *Infrared Physics & Technology* 53 (4) (2010) 267–273.
- [5] J.T. Wang, D.B. Chen, H.Y. Chen, J.Y. Yang, On pedestrian detection and tracking in infrared videos, *Pattern Recognition Letters* 33 (6) (2012) 775–785.
- [6] S.F. Khelifi, M.E. Boudihr, R. Nourine, Compressed telesurveillance video database retrieval using fuzzy classification system, *Lecture Notes in Computer Science* 3656 (2005) 575–584.
- [7] M. Soga, S. Hiratsuka, H. Fukamachi, Y. Ninomiya, Pedestrian detection for a near infrared imaging system, in: *Proceedings of IEEE Conference on Intelligent Transportation Systems*, 2008, pp. 1167–1172.
- [8] C. Fors, S.-O. Lundkvist, Night-time traffic in urban areas: a literature review on road user aspects, Technical Report, VTI rapport 650A, 2009.
- [9] D. Gerónimo, A.M. López, A.D. Sappa, T. Graf, Survey of pedestrian detection for advanced driver assistance systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (7) (2010) 1239–1258.
- [10] P. Dollár, C. Wojek, B. Schiele, P. Perona, Pedestrian detection: an evaluation of the state of the art, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (4) (2012) 743–761.
- [11] L. Zhang, B. Wu, R. Nevatia, Pedestrian detection in infrared images based on local shape features, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [12] H. Sun, C. Wang, B.L. Wang, N. El-Sheimy, Pyramid binary pattern features for real-time pedestrian detection from infrared videos, *Neurocomputing* 74 (5) (2011) 797–804.
- [13] M. Bertozzi, A. Broggi, C.H. Gomez, R.I. Fedriga, G. Vezzoni, M. Del Rose, Pedestrian detection in far infrared images based on the use of probabilistic templates, in: *Proceedings of IEEE Intelligent Vehicles Symposium*, 2007, pp. 327–332.
- [14] Y. Chen, X. Liu, Q. Huang, Real-time detection of rapid moving infrared target on variation background, *Infrared Physics & Technology* 51 (3) (2008) 146–151.
- [15] Y.J. Fang, K. Yamada, Y. Ninomiya, B.K.P. Horn, I. Masaki, A shape-independent method for pedestrian detection with far-infrared images, *IEEE Transactions on Vehicular Technology* 53 (6) (2004) 1679–1697.
- [16] P. Viola, M.J. Jones, D. Snow, Detecting pedestrians using patterns of motion and appearance, in: *Proceedings of International Conference on Computer Vision*, 2003, pp. 734–741.
- [17] C. Papageorgiou, T. Poggio, A trainable system for object detection, *International Journal of Computer Vision* 38 (1) (2000) 15–33.
- [18] Y.B. Zheng, C.H. Shen, R. Hartley, X.S. Huang, Pyramid center-symmetric local binary/trinary patterns for effective pedestrian detection, *Lecture Notes in Computer Science* 6495 (2011) 281–292.
- [19] Y.D. Ma, X.K. Chen, G. Chen, Pedestrian detection and tracking using HOG and oriented-LBP features, *Lecture Notes in Computer Science* 6985 (2011) 176–184.
- [20] P. Sabzmejdani, G. Mori, Detecting pedestrians by learning shapelet features, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [21] B. Wu, R. Nevatia, Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors, *International Journal of Computer Vision* 75 (2) (2007) 247–266.
- [22] A. Miron, B. Besbes, A. Rogozan, S. Ainouz, A. Bensrhair, Intensity self similarity features for pedestrian detection in Far-Infrared images, in: *Proceedings of IEEE Intelligent Vehicles Symposium*, 2012, pp. 1120–1125.
- [23] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.
- [24] Q. Zhu, M.C. Yeh, K.T. Cheng, S. Avidan, Fast human detection using a cascade of histograms of oriented gradients, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 1491–1498.
- [25] Y.W. Pang, Y. Yuan, X.L. Li, J. Pan, Efficient HOG human detection, *Signal Processing* 91 (4) (2011) 773–781.
- [26] T. Watanabe, S. Ito, K. Yokoi, Co-occurrence histograms of oriented gradients for pedestrian detection, *Lecture Notes in Computer Science* 5414 (2009) 37–47.
- [27] Y.W. Xu, X.B. Cao, H. Qiao, An efficient tree classifier ensemble-based approach for pedestrian detection, *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics* 41 (1) (2011) 107–117.
- [28] S. Walk, N. Majer, K. Schindler, B. Schiele, New features and insights for pedestrian detection, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1030–1037.
- [29] C. Wojek, B. Schiele, A performance evaluation of single and multi-feature people detection, *Lecture Notes in Computer Science* 5096 (2008) 82–91.
- [30] M. Enzweiler, D.M. Gavrila, A multilevel mixture-of-experts framework for pedestrian classification, *IEEE Transactions on Image Processing* 20 (10) (2011) 2967–2979.
- [31] J. Gao, Designing real-time multimedia applications on mobile devices, in: *Proceedings of ACM SOSP Workshop on Networking, Systems, and Applications on Mobile Handhelds*, 2011, pp. 1–6.
- [32] S. Maji, A.C. Berg, J. Malik, Classification using intersection kernel support vector machines is efficient, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

- [33] Q. Liu, J.J. Zhuang, S.F. Kong, Detection of pedestrians at night time using learning-based method and head validation, in: *Proceedings of IEEE International Conference on Imaging Systems and Techniques*, 2012, pp. 398–402.
- [34] L.M. Saini, M.K. Soni, Artificial neural network-based peak load forecasting using conjugate gradient methods, *IEEE Transactions on Power Systems* 17 (3) (2002) 907–912.